



Mineria de datos y Patrones

Version 2024-I

Clustering: Mixture of Gaussians

[Capítulo 6]

Dr. José Ramón Iglesias

DSP-ASIC BUILDER GROUP

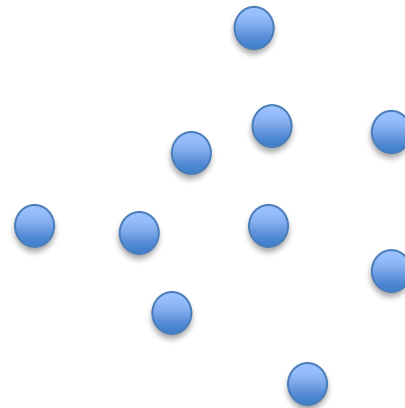
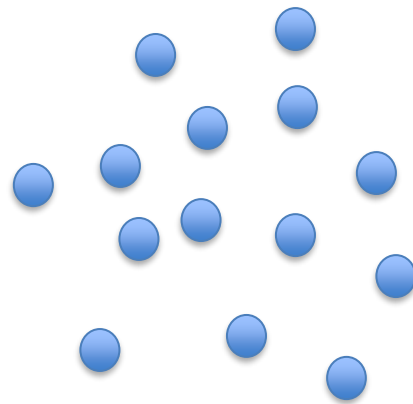
Director Semillero TRIAC

Ingeniería Electrónica

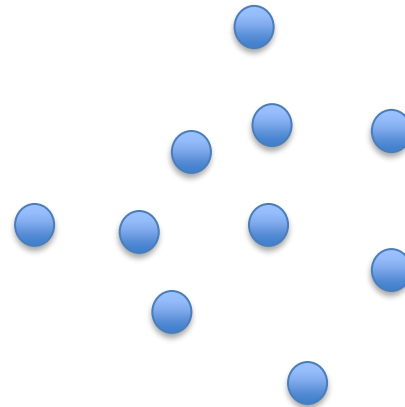
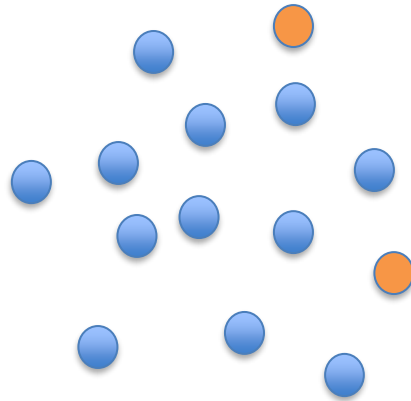
Universidad Popular del Cesar

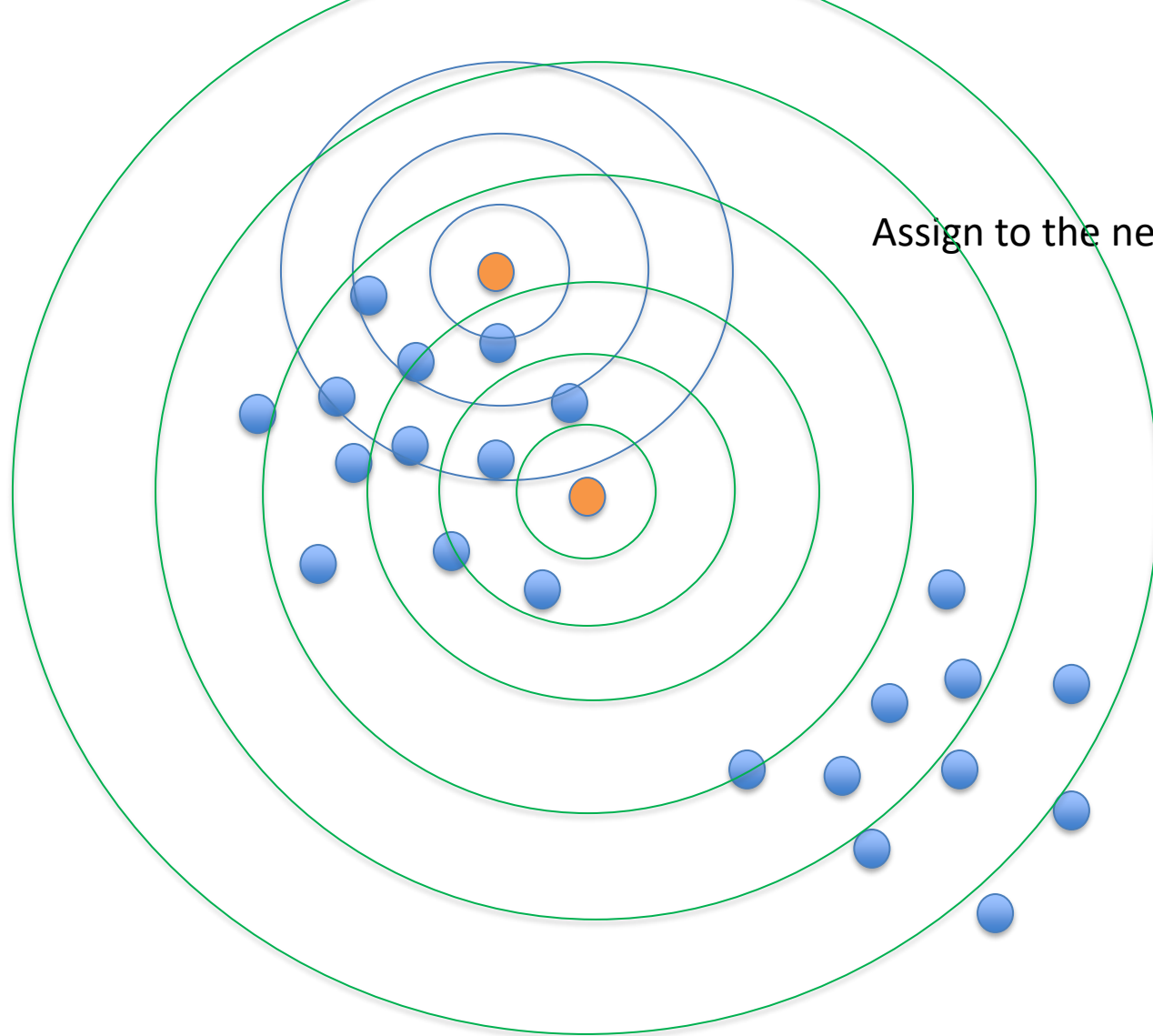
Algorithm:

1. Input Data $X = \{x_1, x_2, \dots, x_N\}$ and number of clusters K
2. Centroids $\{c_1, c_2, \dots, c_K\}$ = random K points of X
3. Initial values for μ_k, Σ_k
4. For each data point x_i
5. Compute Mahalanobis distance $d_{ik} = d(x_i, c_k) \quad i=1, \dots, N, k=1, \dots, K$
6. Assign x_i to the nearest centroid: $y_i = \operatorname{argmin}_j \{d_{ik}\}$
7. Compute for each cluster $\mu_k, \Sigma_k \quad (c_k^* = \mu_k)$
8. if $c_k^* \neq c_k$ then $c_k = c_k^*$ go to step 4
9. Output: $\{c_1^*, c_2^*, \dots, c_K^*\}$ and y_i for $i=1, \dots, N$

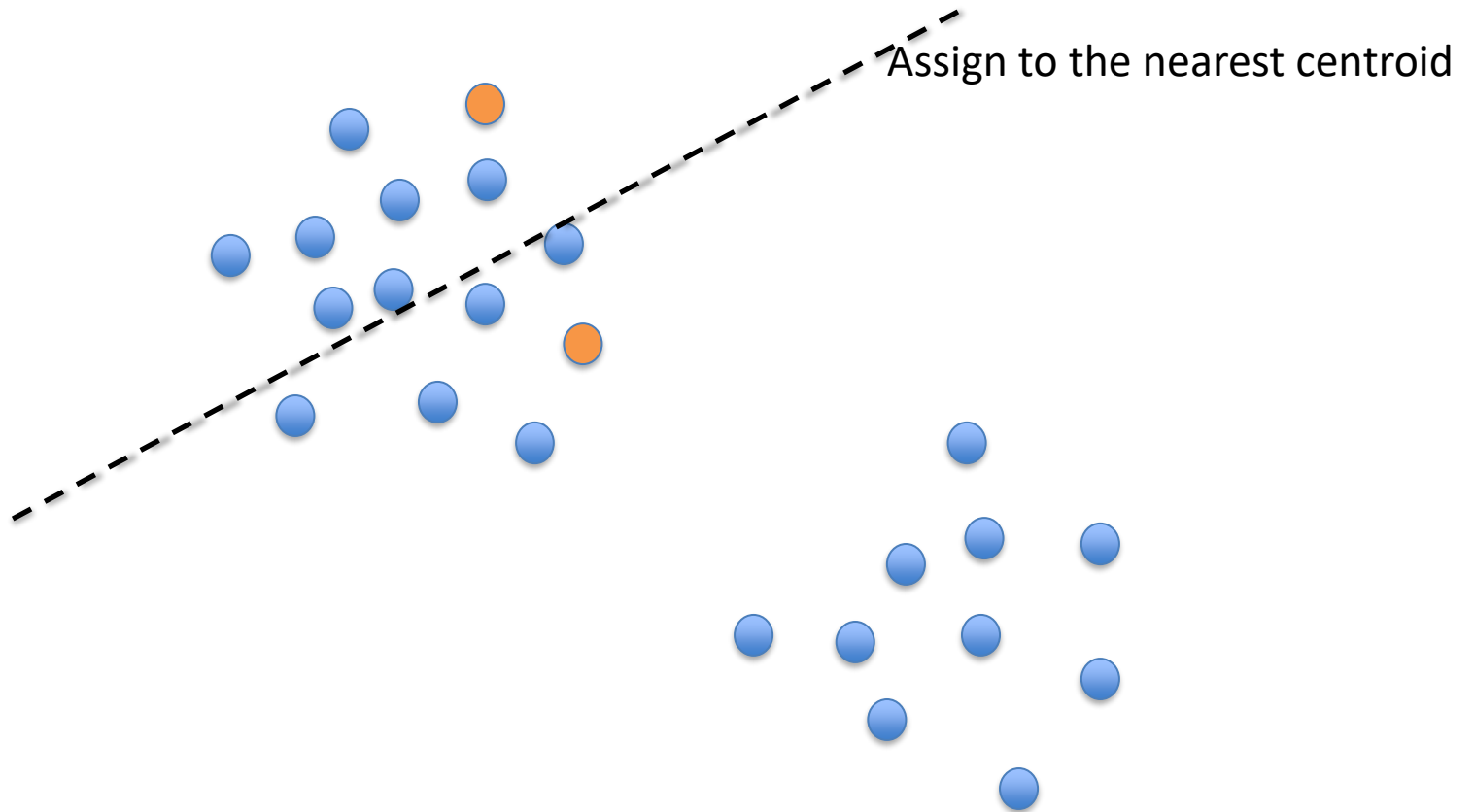


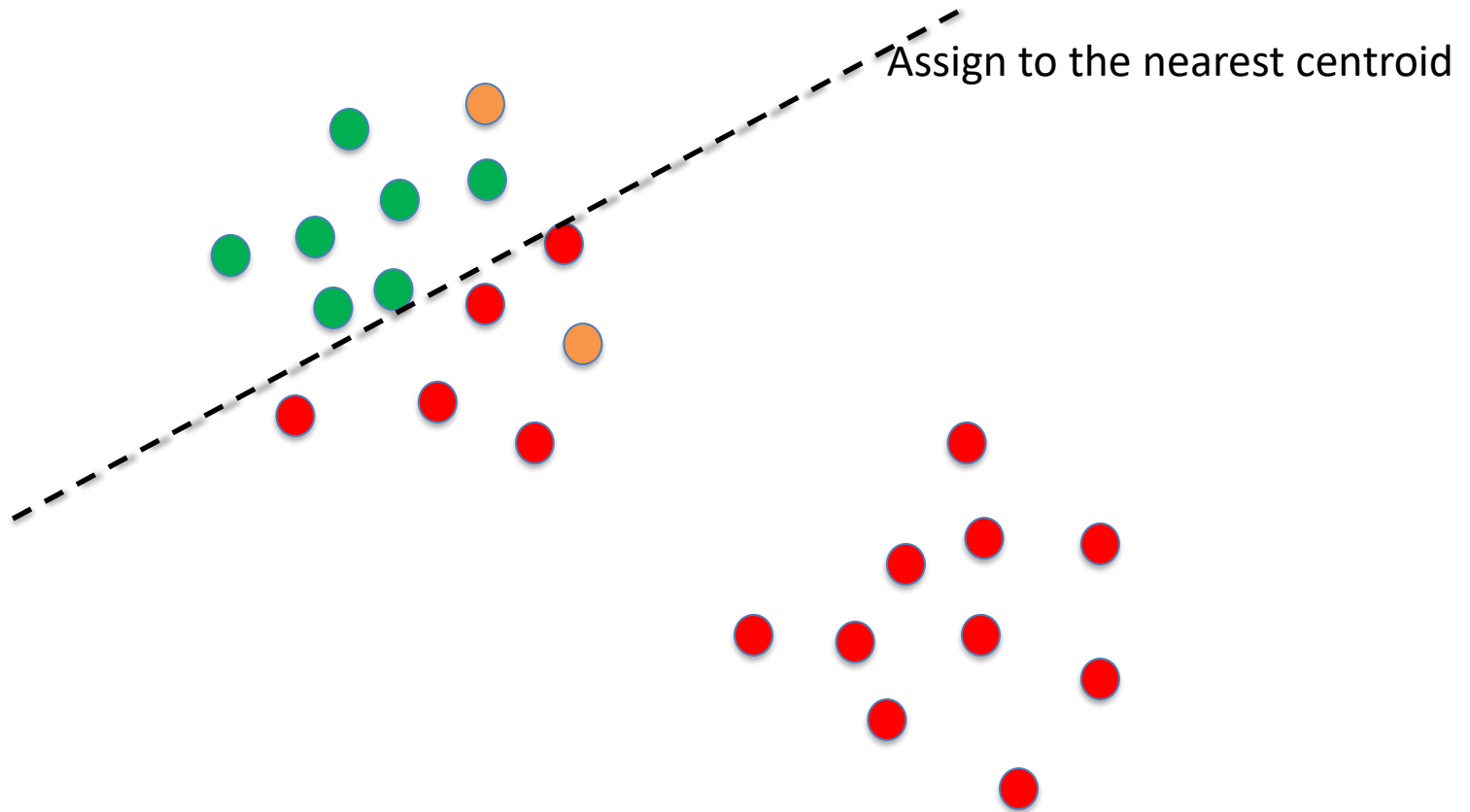
Choose random $K=2$ points (centroids)

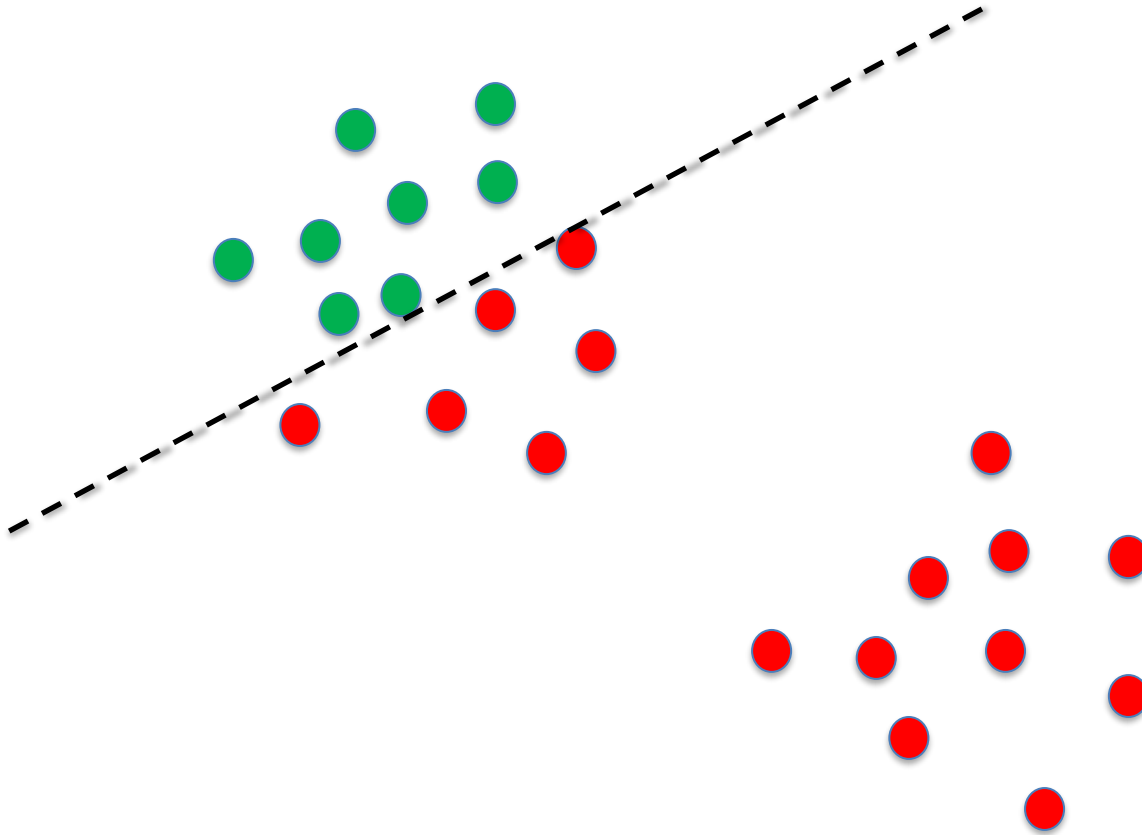




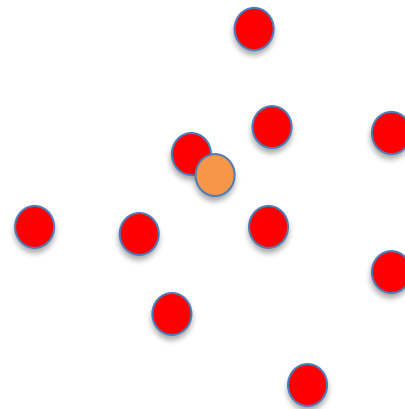
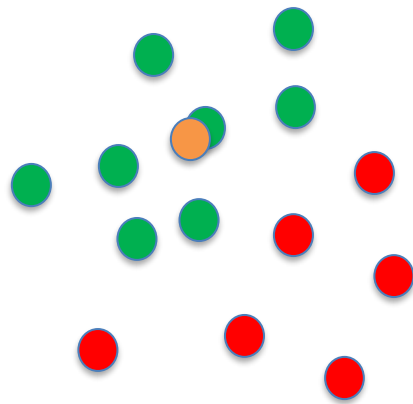
Assign to the nearest centroid

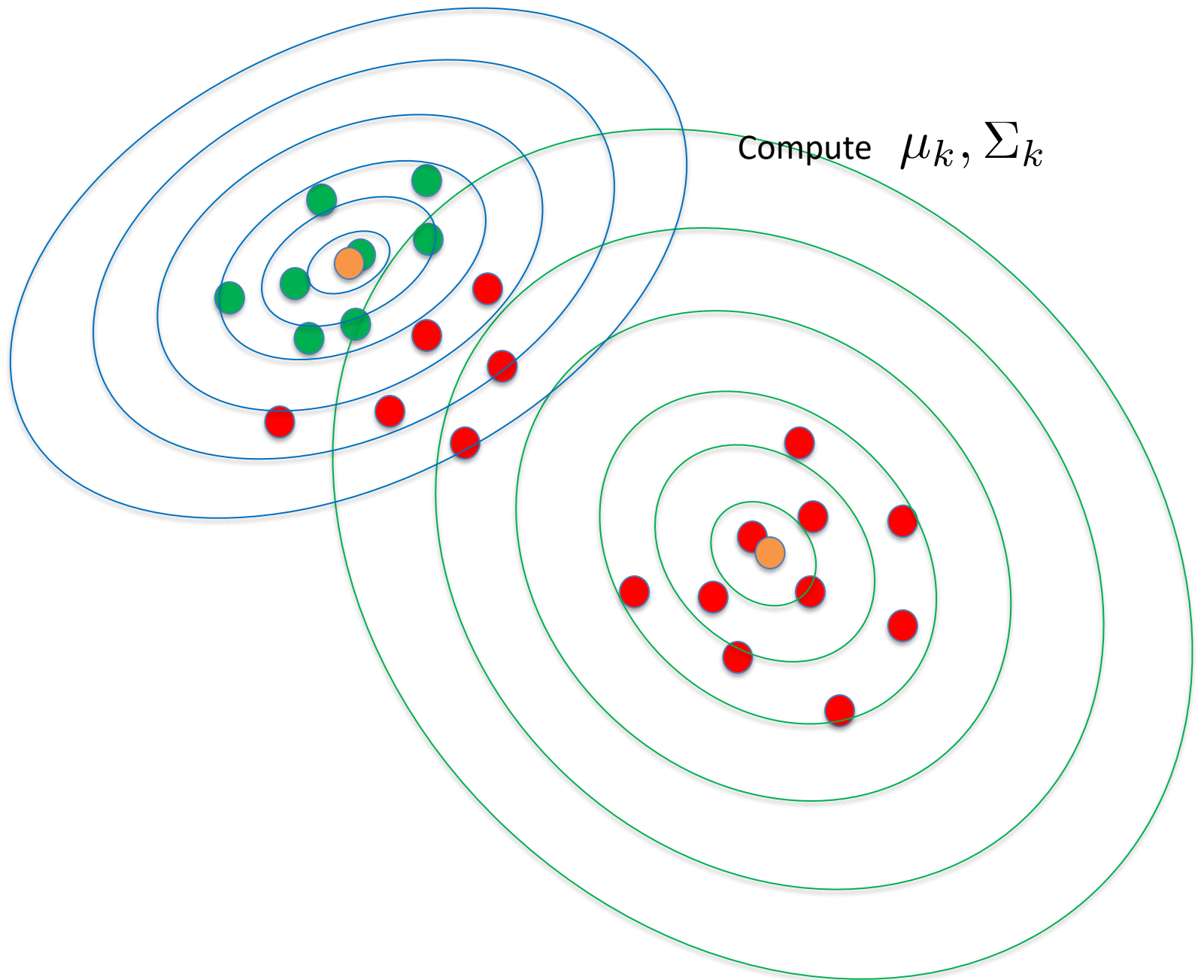


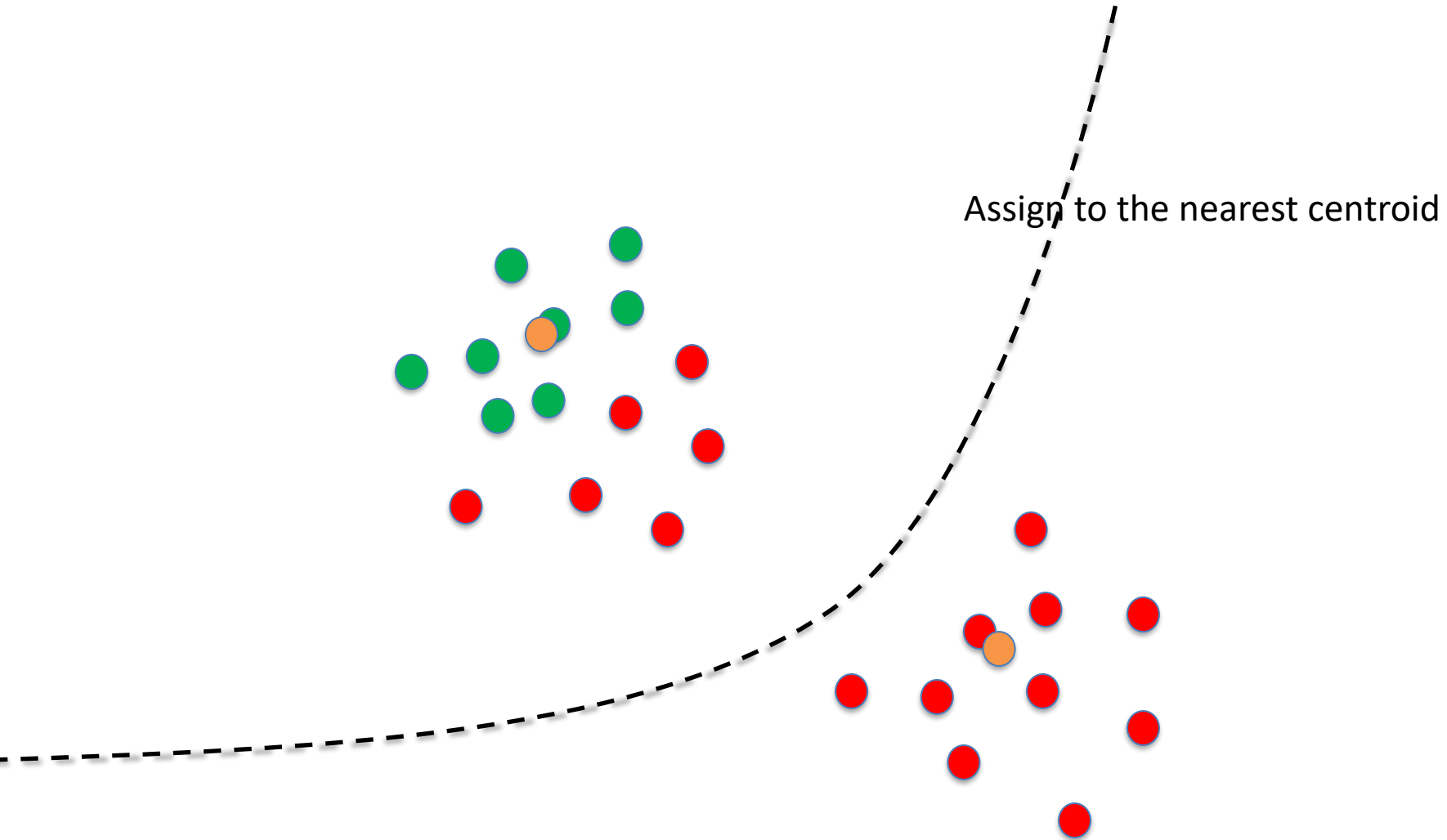


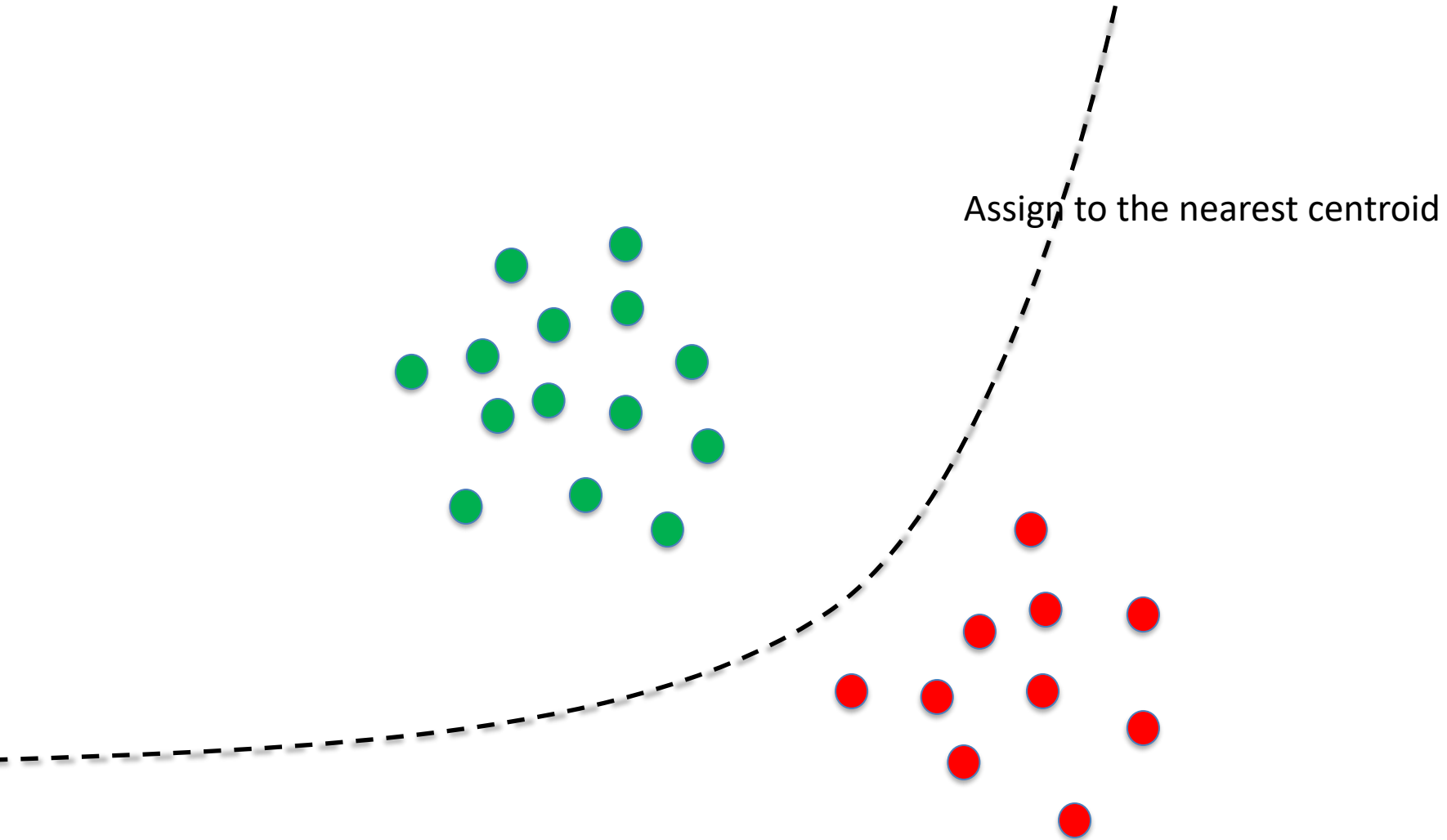


Compute μ_k, Σ_k









Repeat until convergence

